DOCUMENT RESUME

ED 403 780 FL 024 410

AUTHOR Porter, Don; O'Sullivan, Barry
TITLE Writing for a Reader: Does the Nature of the Reader Make a Difference?
PUB DATE 94
NOTE 23p.; Paper presented at the Annual Meeting of the Southeast Asian Ministers of Education Organization Regional Language Center Seminar (29th, Singapore, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Age Differences; *Audience Awareness; College Students; *English (Second Language); Females; Foreign Countries; Higher Education; Interrater Reliability; *Letters (Correspondence); Second Language Instruction; Sex Differences; *Writing Evaluation; *Writing Instruction
IDENTIFIERS *Japan

ABSTRACT

A study investigated how perception of the reader's age in relation to the age of the writer affects assessment of writing. Subjects were 26 Japanese women college students of English as a Second Language, all of whom had recently participated in a home-stay program in an English-speaking country. They were given the task of writing brief letters to three people they knew abroad: a person of approximately their age; an older person; and a younger person. The letter was to discuss the rice shortage in Japan. Four groups of raters included nine each of female and male native speakers of English (NS-F, NS-M) and female and male non-native speakers (NNS-F, NNS-M), all with teaching experience. Each letter was read by three raters from each group. Raters were trained. Each letter was rated according to both holistic and analytic scales. Analysis indicates the ratings varied systematically and significantly with the writer's perceived age of the intended reader. Ratings were highest for older readers, on both scales. A clear tendency was for women raters to score tasks higher than male counterparts on either rating scale. Ratings of NS and NNS were similar. Contains 29 references. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Barry O'Sullivan
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

WRITING FOR A READER: DOES THE NATURE OF THE READER MAKE A DIFFERENCE?

ED 403 780

Don Porter, CALS, University of Reading

and

Barry O'Sullivan, Faculty of Education, Okayama University

1. Introduction

Language is variable. That is to say, the linguistic performance of a language user may vary in characteristic, describable ways with respect to choice of linguistic forms used to convey the message, and with respect to level of formal accuracy in the language deployed. This variation may occur systematically in relation to specifiable linguistic, sociolinguistic and/or psycholinguistic variables. The systematicity of this variation indicates its rule-governed nature.

The above observations have been supported by academic discussion and research over the last thirty years, and are implicitly part of the rationale for communicative approaches to language teaching. Most related research has tended to focus on variability in the Ll; only a limited number of studies, and in recent years, have been expressly devoted to variability in the interlanguage use of L2 learners (Sajjadi, 1994), while a considerable number of studies have turned up evidence on interlanguage variation

FL 024 410

while focusing elsewhere.

It is however clearly of great theoretical and practical interest, both in the teaching of the L2, and in assessing achievement and proficiency in it, for us to understand the ways in which our learners vary in their use of the L2. At a relatively sophisticated level we might wish to know whether our learners are able to vary appropriately in their choice of language in order to achieve diverse purposes, in significantly different situations, and when directing language towards different types of addressee. And at all levels, if language varies in formal accuracy with specifiable socio- and psycholinguistic variables, we would probably wish to know what the variables would be that could be expected to elicit learners' linguistic abilities at their best - or worst. Substantial research is then needed to discover a) what the major relevant sociolinguistic and psycholinguistic variables are for interlanguage, b) how these variables interact, c) the scale of the effects of the variables, and d) the constraints on the variables. For example, the status of the person addressed might have a marked effect on the choice *of* linguistic forms to be used, and thus might be a major relevant sociolinguistic variable. If the gender of the person spoken to were also a major sociolinguistic variable, then a particular combination of gender and status might be markedly potent in affecting the choice of language. The scale of the effect might be such as to make a potentially important difference in the grade awarded in a final examination in the spoken L2. The effect observed might only occur in a specifiable cultural

group. On the other hand, there might be a universal status/gender effect, but exactly how this manifested itself might vary from cultural group to cultural group.

Overwhelmingly, major research relating to variability in interlanguage over the years has concerned spoken interlanguage (Labov, 1972; Selinker, 1972; Bailey Madden & Krashen, 1974; Dickerson, 1974; Larsen-Freeman, 1975; Nakuta, 1976; Tarone, 1979, 1983, 1985; Giles, 1980; Bialystock, 1982; Hyltenstam, 1983; Bell, 1984; Selinker & Douglas, 1985; Bialystock & Sharwood-Smith, 1985; Schachter, 1986; Ellis, 1989; Zuengler, 1989; Gregg, 1990). The literature expressly concerned with determining significant variables affecting L2 writing is relatively sparse, and more often suggests that a given variable is relevant than produces evidence. Variables implicit, suggested or more solidly proposed include topic (e.g. Alderson & Urquhart, 1983; Reid, 1990); genre and rhetorical structure (e.g. Swales, 1990; Connor & Kaplan, 1987); purpose e.g. Witte et al., 1991; Swales, 1990), and lastly audience - or in more everyday parlance, the reader* (e. g. Johns, 1993). Even where the nature of the reader is thought to be a potentially important cluster of variables affecting the writer's performance, there is a tendency for the issue to be discussed in the context of English for Academic Purposes, where the reader is conceived as possessing power over, and having expectations of, the writer.

In contrast, it is our purpose in the rest of this article to present an

empirical investigation into the effects on written language performance of an awareness in the writer of the nature of the intended reader, where the reader has no power over, or significant expectations of, the writer. In so doing, we aim a) to underpin the claim of awareness of reader to being a verifiable variable in interlanguage, b) to provide some evidence of the scale and nature of the effects of the variable, and c) to suggest that the effects of this variable may be culturally constrained.

2. The teaching of writing in Japan

Until recently, writing was rarely taught at second level schools in Japan. In their English classes, students concentrated on the translation of short passages, and on sentence-level, grammar oriented activities. However, from the 1994 academic year, a revised code of study will come into effect. Among the innovations introduced with the new syllabus will be a writing component. The present system of school-based evaluation on student performance - in which the teacher evaluates the student's progress with a series of classroom tests rated on a scale of 1 to 5 - will be maintained. Thus, for the first time, Japanese high school teachers will be required to teach, and to test, students' written work.

While most universities and two-year colleges in Japan offer courses in English, the focus of the majority of these courses is on English literature, with approximately 64% of all English teachers professing to be literature

specialists (Nakane, 1993). The emphasis is on reading texts with the teacher testing the students' understanding by asking them to translate passages and/or write reports in Japanese. Writing-course teachers - often native speakers of English - are expected to evaluate their students' work using grades based on percentage scores.

The disparity in rating methods between these two systems highlights the plight of the average Japanese high school teacher who has had little or no experience or training in teaching or evaluating written work.

3. The hypotheses

It has been frequently observed, informally, that **age** is a culturally significant variable in Japanese society. Japanese society is of course not alone in this respect, and it may even be that age is a significant variable in all societies, affecting social interaction in different ways and to varying degrees.

Nevertheless, Japanese society appears to be particularly marked in this respect. If this observation is correct, then we might expect the interlanguage performance of Japanese learners to reflect a sensitivity to the relative or absolute ages of its users. In fact, if we may be permitted some anecdotal evidence, one of the authors of this article was told by a Japanese audience when discussing the effects of the gender of interlocutors

on the quality of spoken language production, that the gender of an interlocutor would be unlikely to have an effect on a Japanese speaker, but the age of the interlocutor probably would.

If interlanguage sensitivity to age does indeed reflect a true cultural variable, it would seem improbable that the effects of this variable should be restricted to speech. We would expect to find variability with respect to age in the production of the written language, too.

One way in which age-related variability might manifest itself in written language performance might be in "quality of performance", as assessed by the rating scales widely used in the assessment of written language. Such scales normally fall into one of the two categories: "holistic" or "analytic". As both types of scale are used in measuring the same writing skill, we would expect any effects of age-related variability to be perceivable on both types of scale.

Our first and second linked hypotheses were therefore:

> **Hypothesis 1:** Assessment ratings on written language will vary systematically and significantly ($p<.05$) with the perceived age of the intended reader relative to the age of the writer (i.e. younger than the write/ same age as the writer/older than the writer).
>
> **Hypothesis 2:** Hypothesis 1 will be supported on both holistic and analytic rating scales.

Up to this point, we have spoken only of "significant differences" in ratings of writing addressed to readers perceived to be of different ages relative to

the speaker. If differences are indeed found, it is an open question whether they will consist in superior performance being associated with higher-age readers, and inferior performance being associated with lower-age readers. If greater respect is accorded to higher-age readers, this might lead to more care being taken over the writing task, and care in its turn might lead to greater accuracy in syntax and vocabulary use, better organisation of the text, better cohesion, etc. The opposite picture might then be expected to emerge for writing to lower-age readers. Alternatively, the writer might feel more at ease when writing to lower-aged readers, and this ease might translate into greater fluency of writing, wider-ranging and more ambitious use of vocabulary and syntax, and possibly even to greater accuracy. Writing to a higher-age reader might introduce a slight element of stress into the process, and this might translate into over-carefulness, with resulting lower ratings.

The former of these two scenarios seemed to us the more likely. Our third hypothesis was therefore:

> **Hypothesis 3:** Writing to a same-age reader will be rated more highly than writing to a lower-age reader, and writing to a higher-age reader will be rated most highly.

3. Method

3.1. Subjects The subjects for this investigation consisted of a group of twenty-six women students studying English at universities and tertiary

level colleges in Okayama City, Japan. The women ranged in age from 19 to 25 (average age: 20.7), and had taken part in a homestay programme in an English-speaking country in the two-year period preceding the study.

3.2. Procedure The students were given a letter-writing task in which letters were written to three people. As all students had participated in homestay programmes, they were asked to write to real people they had known during their stay abroad. Each student was asked to write to i) a person of approximately the same age as herself (letter S), ii) an older person (letter O), and iii) a person younger than herself. Students were asked to keep their letters short, and to include a topic specified by the investigators. The topic decided on was the rice shortages in Japan, as it was felt that all students would have a fairly clear-cut perspective on this issue, which was highly topical at the time.

In order to avoid the possibility that differences in writing performance might result from the order in which the three letters were written, the order was controlled. Students from each of the three institutions involved in the study were assigned in equal numbers to each of the following orders:

Sequence A:	O	S	Y
Sequence B:	Y	O	S
Sequence C:	S	Y	O

Figure 1 Task sequencing

The tasks were explained and presented to the students by their teachers during classes in their own institutions. Students received an envelope containing the instructions, a return envelope, and the paper on which to write each letter. Also included was a short questionnaire seeking basic demographic information: age, college/university, country visited as part of the homestay programme, duration of stay, and date of the visit. The tasks were performed in the students' own time and were returned to their teacher within one week.

On completion, the resulting 78 letters (3 letters x 26 students) were coded and photocopied. Rater sets were then prepared, each set consisting of three packs of 26 letters. Each of the sets of 26 letters was made up of one letter from every student with O, S and Y as nearly as possible equal in number in each pack. The letters in a pack were arranged in random order and stapled together for raters to score.

3.3 Raters There were four groups of raters: nine male native speakers of English (NS-M) and nine female native speakers (NSF), nine male non-native speakers (NNS-M) and nine female nonnative speakers (NNS-F). All raters had teaching experience, and this ranged from 1 to 17 years (average experience: 6.4 years). Each group of nine raters was given three rater sets of letters to score; thus every letter was read by three people from each rater group, and the load per rater was 26 letters.

All raters were trained, in order to achieve consistency of scoring. Two letters of obviously different levels of writing ability, written in response to the set task but not included in the rater packs, were selected for use as training scripts. These training scripts were first rated by a criterion group of experienced native and non-native teachers, and the ratings suggested by this group were considered to be the "true" ratings for the scripts. The training scripts, together with a set of rater instructions/ were then used to train the four groups of raters described above. When the ratings awarded were within one band of those suggested by the criterion group of raters, it was presumed that a satisfactory level of rater reliability was being achieved (see below for a description of the scales used). Raters whose early judgements were out of step with those of the criterion group were asked first to review the ratings they had awarded and then, if this had proved insufficient, they would have been asked to consult the researchers. No raters needed to take this step.

3.4. Rating Scales: Each letter was rated using two distinctly different scales. The holistic scale was the 1990 revision of the scale used for the Test of Written English (TWE), as presented in Reid (1993: see Appendix 1). This scale permitted raters a range of levels from 1 to 6 or, with half-levels permitted, an eleven-point scale; it allowed raters to give an "overall impression" rating in addition to the more detailed profile yielded by an

analytic scale.

The analytic scale, based on the widely used ESL Composition Profile from Jacobs et al. (1981) called for grades to be assigned on a number of distinct assessment criteria (see Appendix 2). In the original scale, numerical values were awarded on each assessment criterion in a manner which called for a high degree of unguided fine-tuning by the raters. (As an example of the relatively fine distinctions called for, the highest level on the "Content" criterion - "Excellent to Very Good" - still made provision for a range of four marks, from 27 to 30.) On trialling the original scale with native-speaker and non-native speaker volunteer raters, all raters expressed strong reservations about the numerical distinctions. These were consequently changed to a single letter grade to represent each performance level, variable up or down with a simple + or - (see Appendix 3 for the numerical equivalents of the grades used).

4. Data Analysis

The research design provided for independent status on two variables: the linguistic background of the rater (Origin: NNS, NS) and the sex of the rater (Sex: W, M). The third, recurring variable was the relative age of the prospective reader (Letter To: Older, Same Age, Younger). A three-way ANOVA was performed on the data in order to identify main effects, trends, and interactions. This ANOVA was repeated for both the holistic

scale data and the analytic scale data.

In order to identify more precisely the nature of any interactions identified during the initial three-way ANOVA, a one-way ANOVA of the repeated measures from the data provided by both scales was carried out.

Finally, correlations were calculated between the scores achieved on each of the task forms in order to establish the relative validity of the scales.

5. Results

5.1 The analytic scale: The results of the three-way ANOVA on the analytic scale data (see Table 1 below) show significant p-values for the main effects "Sex" and across the repeated measures, although no interactions between the variables were observed. The high p-value for the Origin variable indicates no significant difference between the scores awarded by the NNS and the NS raters.

Source	df	SS	MS	F	p
Origin (A)	1	32.385	32.385	.248	.6199
Sex (B)	1	684.471	684.471	5.233	.0243
AB	1	169.787	169.787	1.298	.2573
Subjects w. groups	100	13080.898	130.809		
Repeated Measure (C)	2	737.374	368.687	10.041	.0001
AC	2	103.572	51.786	1.41	.2465
BC	2	8.807	4.403	.12	.887
ABC	2	.455	.228	.006	.9938
C x Subjs w. groups	200	7343.334	36.717		

Table 1. Three factor repeated measure ANOVA on analytic scale data.

Further analysis of the figures, using the incidence table at Table 2, reveals the pattern of variation in mean scores across the repeated measures, and seems to confirm the tendency in both NNS and NS women raters to award higher scores than male raters. But of particular interest in the current study is the relationship between the score achieved and the age, relative to the writer, of the perceived reader. Letters written to readers perceived to be younger than the writer appear to be scored consistently lower than letters written to readers perceived to be of the same age as the writer, and these in turn are rated lower than letters written to older readers.

	Women Raters			Men Raters			
	Analytic O	Analytic S	Analytic Y	Analytic O	Analytic S	Analytic Y	Totals
NNS Raters	79.09	78.00	75.28	74.74	73.90	70.41	75.24
NS Raters	78.64	75.64	75.60	77.46	74.37	73.59	75.88
Totals	78.86	76.82	75.44	76 10	74 13	72.00	75 56

Table 2. Mean analytic scores for letters to 0, S and Y readers, by NS-M, NS-W, NNS-M and NNS-W raters, each mean based on 26 letters.

A follow-up one-way ANOVA (Table 3) performed on the age-of-reader repeated measure analytic scores confirms that the trends observed in Table 2 are indeed significant. The ANOVA shows significant p-values both between subjects and for the treatments. Moreover, the difference between mean scores for letters to same-age readers and older readers appears to be greater than that between mean scores for letters to same-age readers and younger readers.

Source	df	SS	MS	F	P
Between subjects	103	13987.541	135.607	3.443	.0001
Within subjects	208	8193.542	39.392		
treatments	2	737.374	388.687	10.186	.0001
residual	206	7456.168	38.195		
Total	911	22161.083			

Group	Count	Mean	Std.Dev.	Std.Error
Analytic 0	104	77.481	8.124	.797
Analytic S	104	75.478	8.872	.870
Analytic Y	104	73.718	7.955	.780

Table 3. One-way ANOVA - repeated measures (age of reader) on analytic scale data.

5.2 The holistic scale: It is just possible that effects noted with one type of rating scale may not be apparent with another. In particular, the finer distinctions permitted by analytic assessment may result in greater sensitivity to interlanguage variability in writing than will be found with simpler holistic scales. Therefore, in order to investigate this question, a series of tests similar to those shown above were performed on the data from the holistic scale. Once again, the main effects showed a significant p-value for the sex of the rater, and across the repeated measures (age of reader). As with the analytic scale ANOVA there was no significance in terms of the rater origin, neither was there any indication of cross-variable interaction (Table 4). The trends found in the analytic scale data seem to be confirmed here, although the p-value for the sex-variable indicates an even

stronger significance level than before (.0001, as compared to .0243).

Source	df	SS	MS	F	P
Origin (A)	1 2.837	2.837	2.284	.1339	
Sex (B)	1 24.747	24.747	19.926	.0001	
AB	1 .853	169.787	1.298	.2573	
Subjects w. groups	100 124.198	1.242			
Repeated Measure (C)	2	6.769	3.385	13.530	.0001
AC	2	1.421	.710	2.839	.0608
BC	2	.030	.015	.059	.9423
ABC	2	.096	.048	.192	.8251
C x Subjs w. qroups	200	50.033	.25		

Table 4. Three factor repeated measures ANOVA on holistic scale data.

	Women Raters			Men Raters			
	Holistic O	Holistic S	Holistic Y	Holistic O	Holistic S	Holistic Y	Totals
NNS Raters	4.308	4.167	3.845	3.631	3.468	3.218	3.773
NS Raters	4.412	4.039	4.128	3.908	3.647	3.647	3.963
Totals	4.360	4.103	3.987	3.769	3.558	3.433	3.868

Table 5. Mean holistic scores for letters to 0, S and Y readers, by NS-M, NS-W, NNS-M and NNS-W ratere, each mean based on 26 letters.

The incidence table (Table 5) again confirms a tendency towards higher scoring on letters to older readers, though in this case more notably with the NNS raters. As with the analytic scores, a follow-up one-way ANOVA of the repeated measures (age of reader) was done (Table 6), and this again indicated significance both between subjects and across treatments. Here

again, the difference between mean scores for letters to same-age readers and older readers is greater than that between mean scores for letters to same-age readers and younger readers.

Source	df	SS	MS	F	P
Between subjects	103	152.685	1.482	5.283	.0001
Within subjects	208	58.349	.281		
treatments	2	6.769	3.385	13.518	.0001
residual	206	51.579	.25		
Total	311	210.983			

Group	Count	Mean	Std.Dev.	Std.Error
Holistic 0	104	4.064	.794	.078
Holistic S	104	3.830	.840	.082
Holistic Y	104	3.710	.805	.079

Table 6. One-way ANOVA - repeated measures (age of reader) on holistic scale data.

5.3 Analytic/holistic correlation

The strong relative validity of the analytic and holistic scales is shown in Table 7: for each of the three reader-age groups O, S and Y, the correlations obtained between analytic and holistic ratings are high, ranging from .887 to .912. All other correlations are lower than this, and relatively modest, ranging from a low of .443 to a high of .681.

	Hol.O	Anal.O	Hol.S	Anal.S	Hol.Y
Anal.O	.887				
Hol . S	.681	.595			
Anal . S	.582	.525	.912		
Hol . Y	.551	.523	.630	.488	
Anal . Y	.483	.469	.587	.443	.902

Table 7. OSY holistic/analytic correlation matrix.

5.4 Performance on individual categories on the analytic scale

When the ratings awarded on each of the analytic scale categories are considered in detail (Tables 8, 9, 10, 11 and 12), it will be seen that significant p-values are shown for repeated measures (perceived age of reader) for the categories "content", "organisation" and "vocabulary". It is thus on these categories in particular that awareness of the age of the reader makes itself felt with the Japanese learner. "Language use" (i.e. grammar) and "mechanics" (i.e. spelling, punctuation, etc.) remain non-significant in this respect.

Source	df	SS	MS	F	p
Origin (A)	1	11.643	11.643	1.275	.1616
Sex (B)	1	26.175	26.175	2.866	.0936
AB	1	50.974	50.974	5.582	.0201
Subjects w. groups	100	913.220	9.132		
Repeated Measure (C)	2	79.501	39.751	10.767	.0001
AC	2	15.513	7.756	2.101	.1250
BC	2	1.484	.742	.201	.8181
ABC	2	.448	.224	.061	.9411
C x Subjs w. Groups	200	738.395	3.692		

Table 8. Three factor repeated measures ANOVA on analytic scale category: Content.

Source	df	SS	MS	F	P
Origin (A)	1	11.198	11.198	1.523	.2201
Sex (B)	1	42.733	42.733	5.811	.0178
AB	1	13.695	13.695	1.862	.1754
Subjects w. groups	100	735.409	7.354		
Repeated Measure (C)	2	31.735	15.867	8.734	.0002
AC	2	10.188	5.094	2.804	.0630
BC	2	.134	.067	.037	.9638
ABC	2	1.107	.554	.305	.7~77
C x Subjs w. Groups	200	363.359	1.817		

Table 9. Three factor repeated Measure ANOVA on analytlic scale category: Organisation.

Source	df	SS	MS	F	p
Origin (A)	1	.005	.005	.001	.9746
Sex (B)	1	21.206	21.206	3.990	.0485
AB	1	6 . 456	6 . 456	1. 215	. 2730
Subjects w. groups	100	531.443	5.314		
Repeated Measure (C)	2	48.003	24.001	12.263	.0001
AC	2	1.397	.699	.357	.7003
BC	2	.200	.100	.051	.9503
ABC	2	.819	.409	.209	.8115
C x Subjs w. Groups	200	391.453	1.957		

Table 10. Three factor repeated measures ANOVA on analytic scale category: Vocabulary.

Source	df SS	MS	F	P	
Origin (A)	1	3.976	3.976	.342	.5598
Sex (B)	1	76.487	76.487	6.585	.0118
AB	1	.072	.072	.006	.9374
Subjects w. groups	100	1161.532	11.615		
Repeated Measure (C)	2	18.257	9.128	2.095	.1258
AC	2	5.398	2.699	.619	.5393
BC	2	1.937	.969	.222	.8009
ABC	2	1.820	.910	.209	.8117
C x Subjs w. Groups	200	871.505	4.358		

Table 11. Three factor repeated measures ANOVA on analytic scale category: Language use.

Source	df	SS	MS	F	p
Origin (A) 1 . 494		. 494	. 851	. 3584	
Sex (B) 1 . 676		. 676	1.163	. 2834	
AB	1	. 600	. 600	1. 033	. 3120
Subjects w. groups	100	58.085	.581		
Repeated Measure (C)	2	.192	.096	.616	.5409
AC	2	.330	.165	1.059	.3489
BC	2	.058	.029	.186	.8303
ABC	2	.056	.028	.179	.8362
C x Subis w. groups	200	31.163	.156		

Table 10. Three factor repeated measures ANOVA on analytic scale category: Mechanics

6. Conclusions and discussion.

All three hypotheses were supported: ratings varied systematically and significantly ($p<.05$) with the perceived age of the intended reader; ratings were highest for readers perceived to be older than the writer, and lowest for readers perceived to be younger than the writer; and these were the findings whether holistic or analytic rating scales were used, indicating that any choice between the two methods would have to be made on grounds of practicality alone. The implication here is that, at least with Japanese learners, interlanguage writing varies significantly with the perceived nature of the reader, and that one important feature of that reader, as far as the Japanese EFL writer is concerned, is the perceived age of the readership relative to the writer. Awareness of reader - and particularly awareness of reader's age - is thus clearly an interlanguage variable in Japanese EFL writing.

In addition to the above, a clear tendency was found for women raters to score tasks higher than their male counterparts when using either of the two rating scale types, although the rank ordering of tasks in terms of perceived age of reader remained identical for male and female raters. This finding strongly suggests that the training of raters should have as one of its focuses of attention the elimination of gender differences in the level of marks awarded.

Finally, of particular importance in a testing situation such as that in Japan,

where the latest ministry guidelines for second language teaching call for, among other things, a writing component both "in class" and for college entrance, is the lack of significant difference between scores awarded by non-native and native-speaker raters. While the testing of this component has been seen as problematic, with worries over the capacity of local teachers/ testers to provide reliable results, the findings of this study suggest that the provision of even a minimal training in the use of customised holistic or analytic scale can lead to the awarding by local teachers/ testers of ratings which are very little different from those awarded by English native speaker teachers/testers.

7. Bibliography

Alderson, J.C. and Urquhart, A. 1983: The effects of students' background discipline on comprehension: a pilot study. In A. Hughes and D. Porter (eds) (1983).

Bailey, C., Madden, C. and Krashen, S. 1974: Is there a "natural sequence" in adult second language acquisition? *Language Learning*, 24: 235-244

Bell, A. 1984: Language style as audience design. *Language in Society*, 13: 145-204

Bialystock, E. 1982: On the realtionship between knowing and using linguistic forms. *Applied Linguistics* 3: 181-206

Bialystock, E. and Sharwood-Smith, M. 1985: Interlanguage is not a state of mind. *Applied Linguistics*, 6: 101-117

Connor, U. and Kaplan, R.B. (eds) 1987: *Writinq across languages: analysis of L2 text*. Reading, Massachusetts: Addison-Wesley

Dickerson, L. 1974: Internal and external patterning of phonological variability in the speech of Japanese learners of English: towards a theory of second language acquisition. Doctoral dissertation: University of Illinois.

Ellis, R. 1989: Sources of intra-learner variability in language use and their relationship to second language acquisition. In S. Gass et al. (eds), (1989b).

Gass, S., Madden, C., Preston, D. and Selinker, L. 1989a: *Variation in second language acquisition, Vol.I:* Discourse and Pragmatics. Clevedon, Philadelphia: Multilingual Matters.

Gass, S., Madden, C., Preston, D. and Selinker, L. 1989b: *Variation in second language acquisition, Vol. 2: Psycholinquistic Issues.* Clevedon, Philadelphia: Multilingual Matters.

Giles, H. 1980: Accommodation theory: some new directions. In S. de Silva (ed), (1980).

Gregg, K.R. 1990: The variable competence model of second language acquisition, and why it isn't. *Applied Linguistics*, 11: 364-383

Hakuta, K. 1976: A case study of a Japanese child learning ESL. *Language Learning*, 26: 321-352

Hughes, A. and Porter, D. (eds) 1983: *Current developments in language testing*. New York: Academic Press

Hyltenstam, K. 1983: Data type and second language variability. In H . Ringbom (ed) (1983) .

Jacobs , H . L ., et al . 1981 : *Testing ESL composition: a practical approach*. Rowley, Massachusetts: Newbury House.

Johns, A.M. 1993: Written argumentation for real audiences: suggestions for teacher research and classroom practice. *TESOL Quarterly* 27: 75-90

Kroll, B. (ed) 1990: *Second language writing: research insights for the classroom*. Cambridge: CUP.

Labov, W. 1972: The study of language in its social context. In J.B. Pride and J. Holmes (eds) (1972).

Larsen-Freeman, D. 1975: The acquisition of grammatical morphemes by adult ESL students. *TESOL Quarterly*, 9: 409419

Nakane, C. 1993: *Understanding Japanese Society*. Tokyo: Tuttle

Pride, J. B. and Holmes, J. (eds) 1972: *Sociolinquistics*. London: Penguin Books

Reid, J. 1990: Responding to different topic types: a quantitative analysis from a contrastive rhetoric perspective. In B. Kroll (ed) (1990).

Reid, J. 1993: *Teaching ESL writing*. Englewood Cliffs, New Jersey: Prentice Hall.

Ringbom, H. (ed.) 1983: *Psycholinquistics and foreign language learning*. Abo: Abo Akademi

Sajjadi, S. 1994: Variability in interlanguage. Doctoral thesis, University of Reading.

Schachter, J. 1986: In search of systematicity in interlanguage production. *Studies in Second Language Acquisition*, 8: 119-134

Selinker, L. 1972: Interlanguage. *IRAL*, 10: 209-231

Selinker L. and Douglas, D. 1985: Wrestling with "context" in interlanguage theory. *Applied Linguistics*, 6: 1 90-204